**Proteomics Experiments using HPLC-MS/MS:**

**A Comparison of “Discovery-Driven” and “Hypothesis-Driven” Experiments for peptide measurements.**

**Objectives**

The objective of this laboratory exercise is several-fold:

“Discovery-Driven”: Analyze a semi- complex peptide mixture derived from digested proteins to observe the peptides that are detected by accurate mass and identified by their fragmentation pattern using HPLC coupled to an Orbitrap mass spectrometer.

* Practice and understand peptide ion fragmentation, including the exercise of “de novo” peptide sequencing, in which the fragment ion pattern in the MS/MS spectrum is analyzed to determine the peptide sequence. Raw MS data will be processed using Proteome Discoverer software and MASCOT search engine.

“Hypothesis-Driven”: Select a sub-set of peptides identified in the “Discovery-Driven” experiment and target them in the same sample using HPLC coupled to a triple quadrupole (QqQ) mass spectrometer.

* Optimize the MS/MS fragmentation settings and pick the best fragment ions for targeting on a MS by infusing the sample and recording the best parameters.
* Analyze the data from the “Hypothesis-Driven” experiment to observe and report the average peak area and reproducibility (%CV – coefficient of variation) from replicate measurements of the digested protein sample. Data analysis will be performed using Skyline software (<https://skyline.ms/project/home/software/Skyline/begin.view>)

**Background**

Proteomics is the study of all proteins expressed in a given cell, tissue, or organism. ***“Discovery-driven”*** proteomics is an experimental design in which as many peptides and proteins in the sample are detected by LC-MS/MS and compared across a series of samples with biological differences to see what proteins are changing in abundance. Once several candidate peptides are detected with changes in abundance, these peptides can be screened in a large number of biological samples to precisely determine if the changes are significant, using a ***“hypothesis-driven”*** experiment.

This lab will walk you through how to analyze a sample using discovery-driven proteomics to identify peptides in the sample and the proteins they come from, then to target a small number of peptides using a hypothesis-driven, or targeted approach, observe the peak area over several replicates and calculate the average peak area and %CV.

**Required Materials**

Mobile Phase A: 0.1% formic acid in water (v/v)  
Mobile Phase B: 0.1% formic acid in acetonitrile (v/v)  
4 pmol/uL stock of digested alpha- and beta-casein (in 0.1% formic acid, Sigma)

**Procedure**

**Software download (may be done ahead of time)**

* Skyline Software: <https://skyline.ms/wiki/home/software/Skyline/page.view?name=default>
* Download either 32 or 64-bit version (check your computer’s capabilities)
* For Macs, “Parallels Desktop” is recommended
* The Skyline file with the peptides that will be targeted on the Triple Quadrupole MS will be available on BlackBoard

**Sample Preparation (done ahead of time)**

1. Add 56 uL of Mobile Phase A to a fresh sample vial.
2. Add 4 uL of the 4 pmol/uL stock of **digested alpha- and beta-casein proteins** into the vial with Mobile Phase A and vortex to mix.
3. Aliquot 25 uL each into 2 autosampler vials: 1 for the Orbitrap, the other for the QqQ MS.

**Table 1. Calculate the final concentrations of the sample components and update in the table below**

|  |  |  |  |
| --- | --- | --- | --- |
|  | Starting concentration | Volume added (uL) | Final Concentration  (in 60 uL) |
| Alpha and Beta Casein digest | 4 pmol/uL | 4 |  |

**Question: If 2.5 uL of the sample is injected onto the LC-MS/MS system, what is the final amount in fmol for the following component of the solution being analyzed?**

* + Alpha and Beta Casein digest: \_\_\_\_\_\_\_\_\_\_

**PART I: Discovery-Driven Experimental Parameters for detection of peptides in digested HeLa lysate sample on an Orbitrap MS.**

The first part of this lab is a discovery-driven experiment to determine “what is in my sample?”. A semi-complex sample has been prepared by digesting proteins into peptides using the enzyme trypsin. This first part will use LC-MS/MS to identify as many peptides in the sample as possible and the proteins they come from.

**Table 2. LC Method Parameters**

|  |  |  |
| --- | --- | --- |
| Time (min) | Flow (uL/min) | %B |
| 0 | 25 | 2 |
| 20 | 25 | 30 |
| 25 | 25 | 90 |
| 30 | 25 | 90 |
| 31 | 25 | 2 |
| 50 | 25 | 2 |

Mobile Phase A: 0.1% formic acid in water (v/v)

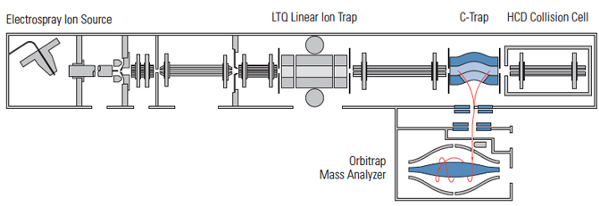
Mobile Phase B: 0.1% formic acid in acetonitrile (v/v)

HPLC column: Acquity UPLC BEH C18, 1 x 150 mm, 1.7 um bead diameter

Injection Volume = 2.5 uL

**MS Method Parameters for Discovery-Driven Experiment; LTQ Orbitrap XL mass spectrometer**

The MS method is set up to operate in “data-dependent” mode. First, the mass spectrometer performs a full scan (MS1 scan), over m/z range 300-2000, to observe the ions detected at that given point in time. Then, the 5 most intense ions are selected for fragmentation and detected in MS/MS mode of analysis. The Orbitrap Mass Analyzer (Figure 1) is used for detection of all precursor ions in the MS1 scan. It provides accurate mass of the precursor ions (operates with <10 ppm mass accuracy). The LTQ Linear Ion Trap is used to fragment and detect the ions in the MS/MS mode of analysis but is not considered an accurate mass analyzer (>100 ppm mass accuracy).



**Figure 1.** A Thermo Fisher Scientific LTQ Orbitrap XL Mass Spectrometer, designed for rapid detection and identification of analytes in samples with accurate mass detection.

The sample prepared consists of a mixture of bovine peptides, generated by the enzyme trypsin, which cleaves at the C-terminus of arginine (R) and lysine (K) residues. There are many tens of peptides in the sample, which makes it very important to use HPLC to separate the mixture over time for MS analysis. Each peptide will have a positive charge on the C-terminus, due to the side chains of arginine and lysine. Since the HPLC conditions are acidic, the N-terminus of each peptide will be charged, resulting in a minimum charge (z) of 2+ per peptide. Based on the other amino acids in the peptide, they may be more highly charged, but the usual range of tryptic peptide charges is 2-5+, which will be observed in the MS1 scan data.

The charge state of the ion can be determined from observing the isotope distribution in the MS1 spectrum (see Figure 2). Due to the contributions of heavy isotopes for each element (13C, 15N, 18O, etc), each peptide ion has an isotope distribution pattern (M+H, M+1, M+2, and M+3, etc) that take into account the natural isotope abundance of all its elements. The spacing between these isotopes is 1 amu, and changes with increasing charge based on the following equation:

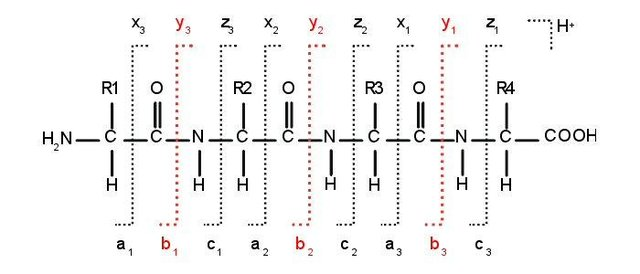
m = 1/z

where m is the spacing between the isotope peaks in m/z units and z is the charge state of the ion. For example, if the isotope peaks are separated by 0.5 m/z units, 0.5 = 1/z, where z = 2, so the ion is doubly charged. Figure 2 contains additional examples of peptides with charge states of 1, 2, 3, and 4, with the m spacing of the isotope peaks shown decreasing with increasing charge.

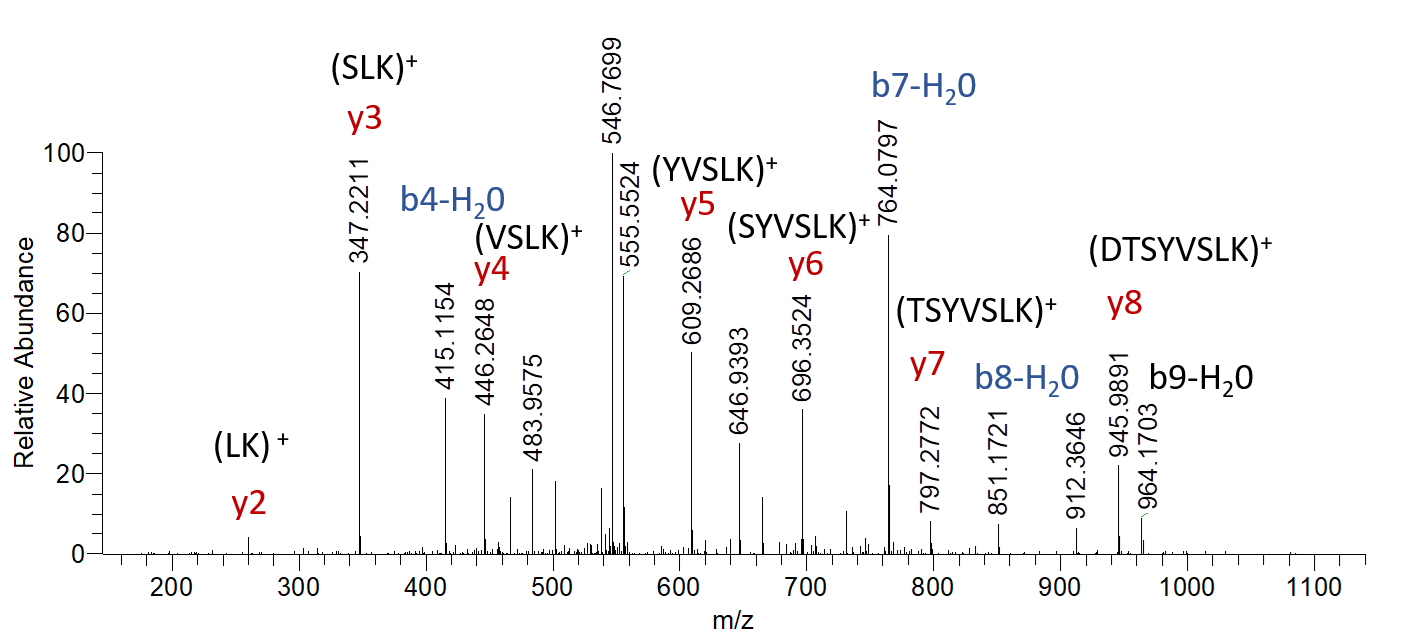


**Figure 2.** Mass Spectra demonstrating how charge state (z) affects the spacing (m) between naturally occurring isotopes in an ion.

Peptide precursor ions detected in the MS1 scan are selected for fragmentation in the mass spectrometer, to break apart the amide bond in the peptide backbone, and to provide fragment ion data to help determine the peptide sequence. Peptides fragment through “Collision Induced Dissociation” (CID) to generate b- and y-ions (see Figure 3). Note the pattern, that if the positive charge is retained on the C-terminal fragment, it is called a ***y-ion***, and if the charge is retained on the N-terminal fragment, it is called a ***b-ion***. The ion’s number is determined by counting from the N-terminus to the fragmentation point for b-ions and from the C-terminus to the fragmentation point for y-ions. An example MS/MS spectrum from a peptide in the MS Qual/Quan QC sample is shown in Figure 4.



**Figure 3.** Peptide Ion Fragmentation Nomenclature



**Figure 4.** MS/MS spectrum with peaks identified by ion and sequence fragment. This is an example of peptide with the sequence ESDTSYVSLK from C-Reactive Protein.

**Investigate LC-MS/MS Data from LTQ Orbitrap XL.**

With your TA in a small group, go through the data-dependent run of the sample and observe the type of data acquired over the chromatographic run.

**Q: Can you explain the difference in the MS1 scan versus the MS/MS (MS2) scan data?**

**How is the information different between each scan type?**

**What information is the MS2 scan data providing?**

**Peptide Sequencing from MS/MS Data.**

With your TA, go over an MS/MS spectrum and calculate the difference in m/z between adjacent fragment ions, then look up these differences to see if they have the same mass as an amino acid (see Amino Acid lookup table in the appendix). Use Figure 4 as an example.

***Complete A2-A5 in Appendix to be turned in with the lab report.***

**Data search against human database to identify peptides and proteins in sample**

A data file from an LC-MS/MS data-dependent run of the sample will be searched in MASCOT and Proteome Discoverer, and a list of proteins and their identified peptides will be provided. The database search processes the data similarly to the manual peptide sequencing exercise you performed, but much, much faster. The result is a list of 100’s of proteins and 1000’s of peptides that have fragment ion patterns that match MS/MS spectra in the data file. This process has become very fast because most protein sequences are known, and the peptides generated during tryptic digestion can be predicted computationally (“*in silico*”). These predictions are compared against the data to find the best match and are reported with a certain level of confidence. This list of peptides and proteins can be used to figure out if any of the peptides should be monitored in a targeted LC-MS/MS method if, for example, they appear to be more abundant in a sick person than in a healthy individual.

**Data Analysis Using Mascot** (converted files will be provided on Blackboard, as well as a table of results from the search)

1. Open analysis file (.raw) in MSConvert
2. Convert file into MGF format (mascot generic file)
3. Open up Chrome and go to matrixscience.com
4. Start a MASCOT MS/MS Ion search.

**Important Parameters:**

Database: Swissprot and Contaminants

Enzyme: Trypsin (2 missed cleavages)

Modifications: Carbamidomethyl (C)

Peptide Tolerance: 1.8 Da

MS/MS tolerance: 0.9 Da

Peptide Charge: +2, +3, +4

Instrument: ESI-Trap

Load your MGF file into your search and “Start Search”

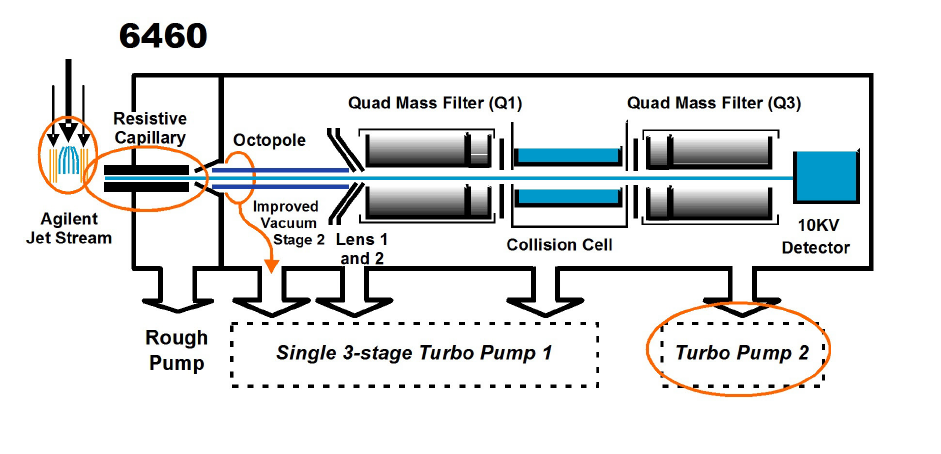
**Q: What protein(s) were identified? What is the sequence coverage?**

**Q: How many peptide matches were identified?**

**Q: How confident are you in your identification?**

**PART II: Hypothesis-Driven Experimental Parameters for detection and quantitation of peptides in digested HeLa lysate sample on a triple quadrupole MS.**

A list of peptides that are deemed biologically interesting have been selected for targeted detection in the same sample so that precise quantitation of these peptides can be determined. Many of these peptides have SIL (stable isotopically labeled, “heavy”) versions spiked into the sample in known quantities so that the “light” versions of the peptides can be quantified. The target list is entered in a Skyline file, and the fragment ions to be monitored are transferred to the MS method on the triple quadrupole MS (Figure 5).



**Figure 5.** An Agilent 6460 triple quadrupole MS, designed for the targeting of analytes with high reproducibility.

**Selection of particular peptides in sample for targeting on the Triple Quadrupole MS**

A subset of the peptides identified from the “Discovery Driven” LC-MS/MS experiment will be selected for a “Hyopthesis Driven” experiment on the triple quadrupole MS. These peptides will be targeted on the triple quadrupole MS using a list of their precursor and product ion fragments.

**Fragmentation (collision induced dissociation, CID) optimization of peptides by infusion on the LTQ XL MS or Orbitrap LTQ XL MS**

The fragmentation products of peptides in the sample were observed on the LTQ Orbitrap MS, but in order to improve sensitivity for their detection on the Triple Quadrupole MS, we will infuse a concentrated stock solution and optimize the collision energy.

* Obtain a concentrated stock solution of alpha and beta casein in 50% acetonitrile/0.1% formic acid.
* Fill a glass syringe with 500 uL and set up on the syringe pump on the LTQ XL MS or Orbitrap LTQ XL with a flow rate of 5 uL/min
* Set the MS to do a “full scan” and observe the precursor peptide ions, and their charge states, if applicable
* Select one of the precursors to fragment and write down the m/z to the nearest 0.1 m/z unit
* Change the MS scan type to perform “MS2” and input the precursor m/z with a CE (collision energy) of 0.
* Set the scan range to 150-2000 m/z
* Apply changes. The spectrum should now contain only the peptide m/z signal you selected.
* Adjust the CE in 5 V increments, hitting “apply” after each setting. Keep increasing the CE voltage until the precursor m/z is a minor peak in the spectrum and the product ions are dominant.
* Write down the m/z of the 10 most intense product ions observed in the spectrum and the final CE setting you used.

**Targeting Peptides on the Triple Quadrupole MS using LC Separation**

**Procedure**

1. Open the Skyline file (MSQC\_Casein\_MRM\_template.sky) provided for this experiment and check that the list of peptides in the file match the ones listed in the method for the triple quadrupole MS.
2. Set up the LC gradient with the following conditions (the same column, solvents, and sample will be used), see Table 4.
3. Inject 3 replicates of the sample so that reproducibility for peak area, peak area ratio, and retention time can be determined.

**Table 4. LC gradient conditions for hypothesis-driven experiment**.

|  |  |  |
| --- | --- | --- |
| Time (min) | Flow (uL/min) | %B |
| 0 | 25 | 2 |
| 20 | 25 | 30 |
| 25 | 25 | 90 |
| 30 | 25 | 90 |
| 31 | 25 | 2 |
| 50 | 25 | 2 |

**Analysis of data in Skyline**

1. Once each file is acquired, it can be imported into Skyline for analysis. Go to File/Import/Results; select “OK” from the default option listed and select the data file for import.
2. Click on each peptide sequence to observe the peak that has been detected.
3. After importing replicates 2 and 3, ensure the replicates for each peptide have the same approximate retention time (ie, the same peak is selected). Adjust the peak selection if necessary.
4. Report the average and %CV (coefficient of variation) in peak area from the 3 replicate injections for the peptides in the Skyline file. Use Table 5

**Table 5: Peak Area Reproducibility for Peptides 1-9**

|  |  |  |
| --- | --- | --- |
| Peptide | Average Peak Area | Peak Area %CV |
| EAMAPK |  |  |
| EMPFPK |  |  |
| VLPVPQK |  |  |
| AVPYPQR |  |  |
| DMPIQAFLLYQEPVLGPVR |  |  |
| HQGLPQEVLNENLLR |  |  |
| FFVAPFPEVFGK |  |  |
| VNELSK |  |  |
| YLGYLEQLLR |  |  |

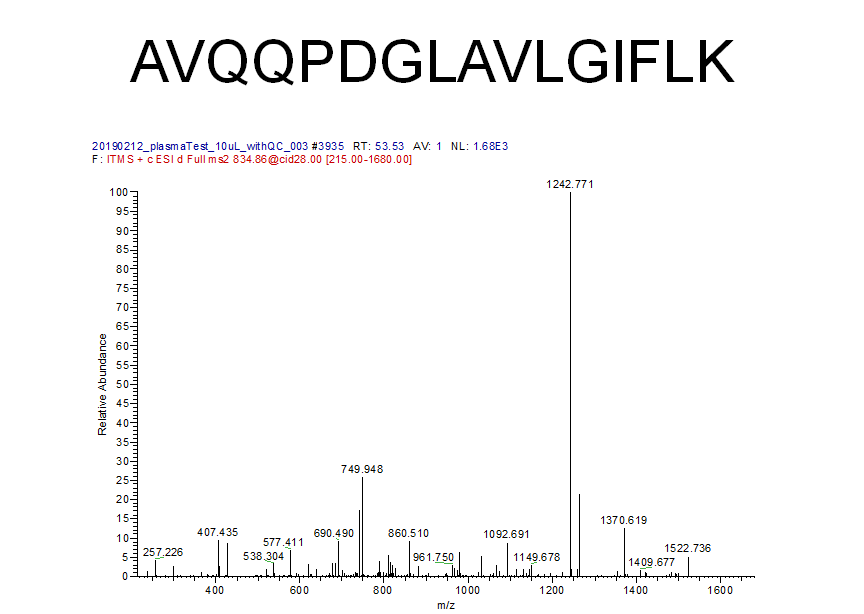
**Appendix**

A1. Amino Acid Lookup Table.

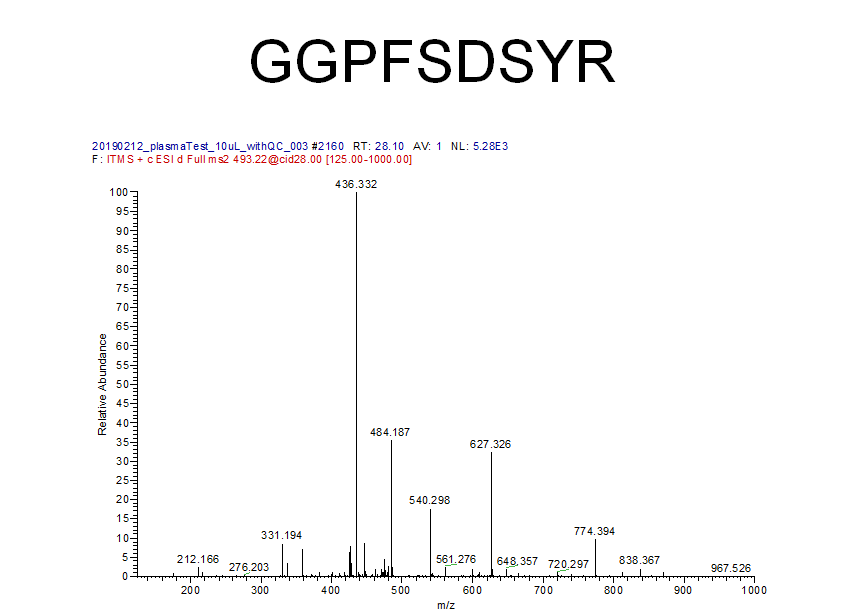
Use the MW from the amino acids below to see if any of them match the difference in m/z from fragment ions in the MS2 spectra in Appendix A2-A5. You may also use the website: <http://prospector.ucsf.edu/prospector/cgi-bin/msform.cgi?form=msproduct> to type in the peptide sequence (for spectra in Appendix A2-A4) to help match up fragment ions with their identity (ie, b2+, y6+, y8++)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Amino Acid | 3 Letter Code | 1 Letter Code | Monoisotopic Mass | Average Mass | Elemental composition |
| Glycine | Gly | G | 57.02146 | 57.05 | C2H3NO |
| Alanine | Ala | A | 71.03711 | 71.08 | C3H5NO |
| Serine | Ser | S | 87.03203 | 87.08 | C3H5NO2 |
| Proline | Pro | P | 97.05276 | 97.12 | C5H7NO |
| Valine | Val | V | 99.06841 | 99.13 | C5H9NO |
| Threonine | Thr | T | 101.04768 | 101.1 | C4H7NO2 |
| Cysteine | Cys | C | 103.00919 | 103.1 | C3H5NOS |
| Isoleucine | Ile | I | 113.08406 | 113.2 | C6H11NO |
| Leucine | Leu | L | 113.08406 | 113.2 | C6H11NO |
| Asparagine | Asn | N | 114.04293 | 114.1 | C4H6N2O2 |
| Aspartic Acid | Asp | D | 115.02694 | 115.1 | C4H5NO3 |
| Glutamine | Gln | Q | 128.05858 | 128.1 | C5H8N2O2 |
| Lysine | Lys | K | 128.09496 | 128.2 | C6H12N2O |
| Glutamic Acid | Glu | E | 129.04259 | 129.1 | C5H7NO3 |
| Methionine | Met | M | 131.04049 | 131.2 | C5H9NOS |
| Histidine | His | H | 137.05891 | 137.1 | C6H7N3O |
| Phenylalanine | Phe | F | 147.06841 | 147.2 | C9H9NO |
| Arginine | Arg | R | 156.10111 | 156.2 | C6H12N4O |
| Tyrosine | Tyr | Y | 163.06333 | 163.2 | C9H9NO2 |
| Tryptophan | Trp | W | 186.07931 | 186.2 | C11H10N2O |

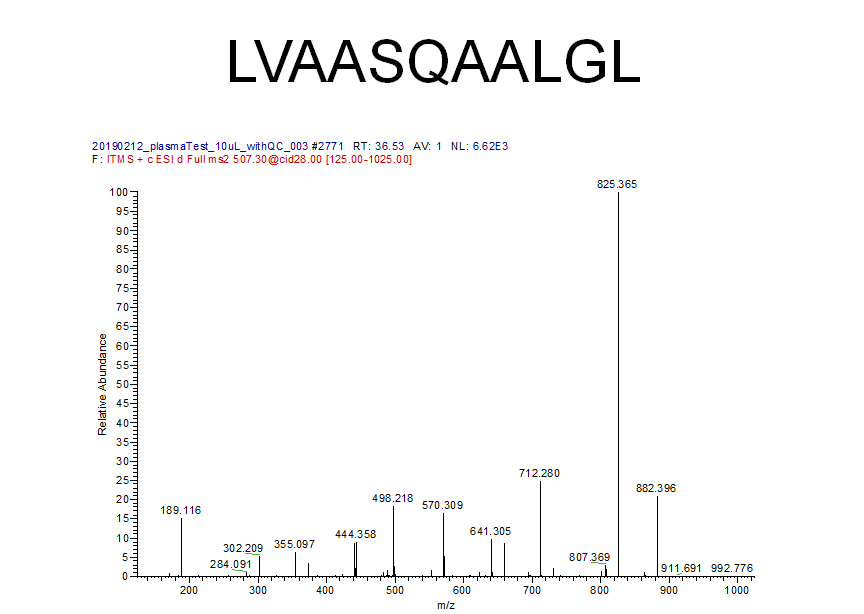
**A2. MS/MS Spectrum 1. Label the fragment ion peaks with the proper fragment identity (ie, b2+, y6+, y8++, etc)**



**A3. MS/MS Spectrum 2. Label the fragment ion peaks with the proper fragment identity (ie, b2+, y6+, y8++, etc)**



**A4. MS/MS Spectrum 3. Label the fragment ion peaks with the proper fragment identity (ie, b2+, y6+, y8++, etc)**



**A5. MS/MS Spectrum 4 (BONUS CREDIT!!). Label the fragment ion peaks with the proper fragment identity (ie, b2+, y6+, y8++, etc). This spectrum does not have a peptide sequence identified, use the peptide lookup table in A1 to help you identify loss of amino acids and try to piece the peptide sequence together.**

290.14490

601.36755

273.11868

389.21295

175.11858

502.29742

72.08118

100

200

300

400

500

600

700

m/z

0

100

200

300

400

500

600

700

I

n

t

e

n

s

i

t

y

[

c

o

u

n

t

s

]

(

1

0

^

3

)

FTMS, HCD@28.00, z=+2, Mono m/z=350.72049 Da, MH+=700.43370 Da, Match Tol.=20 ppm

**FINAL QUESTIONS:**

1. Explain the qualitative differences between the chromatograms observed from the discovery-driven and hypothesis-driven experiments.
2. Why can the targeted experiments be run a bit faster than the discovery-driven experiments?
3. Draw a picture of a nanocolumn tip and what the stationary phase looks like while it is packing into the column.